Systematic Literature Review: Hybrid Intelligence Frameworks in Detecting and Mitigating Social Engineering Attacks

Ani Ukamaka Eucharia[1] **Omeje Kingsley Nnabuike**[2] **Ijegalu Lilian Njideka**[3]

Orcid ID:0000-0001-8986-6561

Ngene Samuel [4] **Anikwe-Onyiwa Chioma Virgina**[5]

Onu Sunday[6] **Ori Silas Ene**[7]

Orcid ID:0009-0005-1216-3135 ***Orcid ID:0009-0009-4063-2639***

1-7, Department of Computer Science, David Umahi Federal University of Health Sciences, Uburu

Abstract

Social engineering attacks pose significant cybersecurity threats, exploiting human vulnerabilities instead of relying solely on technical flaws. This systematic literature review aims to examine hybrid intelligence frameworks that leverage advanced technologies alongside human factors to better detect and mitigate these attacks in dynamic environments. Following the PRISMA guidelines, we identified and analyzed relevant studies published between 2016 and 2024. The review categorizes the literature into three primary themes: integration of technology and adaptive measures in detection methodologies, the impact of user behavior on vulnerability, and proposed frameworks or models for both prevention and response strategies.

Key findings indicate that hybrid models incorporating machine learning and behavioral analysis significantly enhance detection capabilities, paving the way for proactive mitigative actions. Mouton et al. outlined detection models applicable to social engineering, emphasizing the need for ongoing adaptation in security frameworks [1]. [2]. highlighted the critical role of user awareness and behavior in mitigating risks, advocating for a psychologically informed approach to social engineering defense [3] . Recent approaches leveraging generative AI illustrate the evolving landscape of social engineering, showcasing automated and highly targeted attacks [4].

Overall, this review identifies significant gaps in traditional defensive mechanisms, underlining the necessity for a multidisciplinary perspective that combines behavioral and technological insights to enhance resilience against social engineering threats. This study lays the groundwork for future inquiries into hybrid frameworks, aiming for a more comprehensive understanding of their roles in cyber security.

Keywords: Hybrid, Intelligence, Framework, Detecting, Mitigating, Social Engineering

Introduction

Social Engineering Attacks (SEAs) represent one of the most persistent and manipulative forms of cybersecurity threats, exploiting human psychology rather than technical vulnerabilities. As attackers grow more sophisticated, traditional rule-based systems struggle to keep pace. Hybrid Intelligence Frameworks combining machine intelligence (like ML, NLP, DL) with human oversight or reasoning have shown promise in countering these evolving threats. This SLR uses the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology to identify, assess, and synthesize literature related to hybrid intelligence systems in SEA detection and mitigation. This introduction contextualizes the growing threat of SEAs, the limitations of traditional mitigation techniques, and the emergent role of hybrid systems in this security arms race.

Social engineering constitutes an attack vector that exploits human affectivity, bypassing traditional cybersecurity measures designed to protect technological infrastructures. Given the rapid growth of digital interactions, contemporary cybersecurity has seen an increase in social engineering tactics, ranging from phishing to impersonation, and manipulation. Addressing the challenges posed by these deceptive strategies requires innovative and adaptable defenses.

Social engineering attacks capitalize on psychological triggers, addressing the human element rather than attempting to breach technological barriers. High-profile incidents like the Target breach (2013) and the social engineering ploys employed against companies like Twitter highlight the need for more robust defenses that address both technological and human vulnerabilities [5]; [6]. This trend has propelled research into multifaceted approaches that blend artificial intelligence with human behavioral insights, heralding the rationale for hybrid intelligence frameworks.

Existing literature recognizes hybrid intelligence frameworks—integrative systems utilizing both advanced technology and cognitive psychological principles—as crucial in effectively tackling social engineering attacks. Hybrid frameworks typically combine machine learning algorithms with human-centric design to enable systems that not only react to but also anticipate malicious behaviors, thereby evolving current preventive strategies [1]; [5]
At its essence, social engineering leverages psychological principles to influence human behavior (Wang et al., 2021). Attackers study personality traits, observe body language, and apply manipulation techniques that exploit trust, authority, and urgency. The success of these attacks is largely due to their ability to exploit inherent cognitive biases and psychological vulnerabilities in humans, such as the tendency to comply with authority or to respond reflexively under pressure [8]. These cognitive weaknesses form the basis for core social engineering strategies, which revolve around exploiting trust and human error rather than technical flaws. Traditional tactics like **pretexting**, **baiting**, **quid pro quo**, and **tailgating** have evolved beyond basic deception into more complex, blended campaigns. For instance, attackers may combine email phishing with social media reconnaissance and follow-up phone-based vishing to enhance credibility and increase the likelihood of success. As Wang et al. (2021) suggest, these multi-channel strategies reflect the adaptability and ingenuity of modern attackers, making detection increasingly difficult. Consequently, the dynamic nature of social engineering demands an integrated defense approach that considers both technological safeguards and the psychological dimensions of user behavior. The digital age has introduced a complex interplay between technological advancement and cybersecurity threats, Among the most pressing concerns are social engineering and ransomware attacks, which are increasingly being amplified and transformed by AI [9]. In recent years, there has been a marked increase in both the prevalence and sophistication of these attacks, driven in large part by emerging technologies such as artificial intelligence (AI), deepfakes, and the pervasive use of social media. Generative AI, in particular, has drastically reshaped the threat landscape by enabling attackers to craft highly convincing, personalized, and scalable campaigns [10]. These tools allow adversaries to automate social engineering techniques, such as phishing emails, synthetic voice calls (vishing), and social media impersonation with unprecedented precision and realism. As [13] note, the constant evolution of social engineering tactics, powered by AI, makes them more effective at bypassing traditional defenses and manipulating human targets. Real-world incidents, including business email compromise and AI-enhanced phishing,

continue to highlight the devastating impacts of such attacks across sectors like finance, healthcare, education, and critical infrastructure. In response, the research community has intensified its focus on understanding and mitigating these threats. Detection methods have progressed from basic rule-based systems and awareness training to sophisticated approaches using machine learning (ML), behavioral analytics, and hybrid frameworks. However, while these advances are notable, user awareness alone has proven insufficient, particularly in the face of context-aware and adaptive attack strategies. Moreover, the current body of research remains fragmented: many studies isolate specific attack types, most commonly phishing, without addressing the broader spectrum of social engineering strategies. Detection techniques also vary significantly in methodology, scope, and evaluation metrics, making cross-comparison and standardization difficult. As attackers continue to evolve with the aid of AI-generated text, deepfake audio, and the proliferation of connected devices, many existing detection solutions risk becoming obsolete. These challenges emphasize the urgent need for a comprehensive synthesis of recent scholarly efforts. This systematic literature review, therefore, aims to consolidate findings from 2020 to 2025, map the current research landscape, identify gaps, and propose informed directions for future research and defense strategies in the fight against social engineering.

Categorization of Social Engineering Attacks

Social engineering attacks can be broadly categorized based on their primary delivery method and the psychological manipulation tactics employed. These categories often overlap, as attackers may combine multiple techniques for greater effectiveness.

A. Digital/Remote Social Engineering Attacks

These attacks primarily occur through digital communication channels, without direct physical interaction.

1. **Phishing**: According to [11], Phishing is the most prevalent form of social engineering, where attackers send fraudulent communications (e.g., emails, messages) appearing to come from a legitimate and trustworthy source to trick recipients into revealing sensitive information or clicking malicious links. Phishing attacks can be highly personalized and automated, especially with the advent of generative AI [12].

- **Email Phishing**: According to [13],_ Email phishing is the most common type, involving deceptive emails. These emails often contain links to fake websites or attachments that, when clicked or downloaded, can install malware or steal personal information [13].
- **Spear Phishing**: A highly targeted form of phishing aimed at specific individuals or organizations [14]. Attackers conduct extensive research to create personalized and convincing messages, leveraging the victim's name, job title, or other known details [13].
- **Whaling**: A specialized spear phishing attack that targets high-profile individuals, such as executives or senior management [15]. The objective is often to obtain sensitive corporate information or authorize significant financial transactions.
- **Vishing (Voice Phishing)**: Involves attackers making phone calls, impersonating legitimate entities (e.g., banks, government agencies), to trick victims into disclosing personal or financial information over the phone [11].
- **Smishing (SMS Phishing)**: This form of phishing is conducted via text messages (SMS), it contains malicious links or instructions to call fraudulent phone numbers [13].
- **Angler Phishing**: [13] explains Angler Phishing as an exploit perpetrated on social media platforms by creating fake profiles or sending phishing messages through social media channels to deceive users into divulging personal information.
- **HTTPS Phishing**: involves creation of fake websites using HTTPS protocol to appear legitimate and secure, tricking users into entering sensitive data [13].
- **Clone Phishing**: Attackers create a replica of a legitimate email with slight modifications, such as replacing a genuine link with a malicious one [13].
- **Pharming**: A more technical variant where attackers manipulate DNS settings or use malware to redirect users from legitimate websites to fake ones without their knowledge [13].

- **Pop-up Phishing**: Utilizes fake pop-up windows that mimic security warnings or software updates to lead users to phishing websites or malware installation [13].

2. **Pretexting**: This involves creating a fabricated scenario or "pretext" to gain the victim's trust and extract information [16]. The attacker assumes a persona (e.g., an IT support technician, a bank representative) and crafts a believable story to manipulate the target into providing information or access [17]. For example, an attacker might call an employee pretending to be from IT support, claiming there's a critical system issue requiring immediate login credentials.

3. **Baiting**: This technique involves offering something enticing (the "bait") to the victim in exchange for sensitive information or to lead them into a trap [18].Baiting can come in 2 different forms; Physical Baiting which involves leaving malware-infected physical devices (e.g., USB drives) in public places, hoping someone will pick them up and plug them into their computer while Digital Baiting Could involve enticing offers online, such as free downloads, movies, or music, which are actually Trojan horses or lead to malicious websites [19].

4. **Quid Pro Quo**: An attacker offers a service or benefit in exchange for information or action [18]. Unlike baiting, which offers something tangible and immediate, quid pro quo usually involves a service. For example, an attacker might call random numbers in a company, claiming to be technical support offering to fix a problem, and then ask for credentials to "assist" them.

5. **Business Email Compromise (BEC):** BEC is a sophisticated cybercrime that leverages social engineering to defraud organizations, often resulting in significant financial losses [20]. These attacks typically involve impersonating a trusted entity, such as a CEO or a vendor, to induce victims to transfer funds or sensitive data to the attacker's control

B. Physical/In-Person Social Engineering Attacks

These attacks require the attacker to be physically present or to directly interact with the victim in person.

1. **Tailgating (Piggybacking)**: An unauthorized person follows an authorized person into a restricted area [16]. This often occurs when someone holds a door open for another person, who may appear to be an employee but lacks proper credentials. For example, an attacker might carry a box to appear busy and legitimate, relying on an employee's courtesy to hold a secured door open for them.

2. **Shoulder Surfing**: An attacker directly observes a person entering sensitive information (e.g., passwords, PINs) from a close distance [19]. This can occur in public places like cafes, airports, or even within an office environment. For example, an attacker might stand behind someone at an ATM or while they are using a laptop in a public space to discreetly watch their input.

3. **Dumpster Diving**: Searching through discarded materials (e.g., trash, recycling bins) to find valuable information that can be used in a social engineering attack [10]. This information can include documents with sensitive data, old hardware, or even sticky notes with passwords. For example, discarded financial statements, employee directories, or technical manuals can provide attackers with crucial details for other social engineering tactics.

4. **Reverse Social Engineering**: In this less common but highly effective method, the victim approaches the attacker for help. The attacker first creates a problem (or makes the victim believe there's one) and then positions themselves as the only one who can solve it. This builds immense trust, making the victim more likely to disclose information [19]. For example, an attacker might intentionally disable an employee's computer, then wait for the employee to seek "IT help," where the attacker then offers assistance and obtains credentials.

Research Questions

RQ1: What hybrid intelligence frameworks have been proposed for detecting or mitigating SE attacks?

RQ2: What techniques (e.g., ML, NLP, game theory) are most common in these frameworks?

RQ3: What are the research gaps and challenges in implementing these frameworks?

Methodology

PRISMA Framework

The methodology for this systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) framework, encompassing key stages—identification, screening, eligibility, and inclusion.

1. **Identification**: An extensive search was performed using academic databases such as IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar, along with various cybersecurity journals. The search was restricted to studies published from 2016 to 2024 and included keywords such as "hybrid intelligence," "social engineering," "cybersecurity," "machine learning," and "behavioral analysis." The identification process yielded a total of 750 records.
2. **Screening**: After the initial assessment, duplicates were removed, resulting in 620 unique records. Subsequently, titles and abstracts of these records were screened to determine relevance to the research questions. Studies that focused on hybrid intelligence frameworks, detection, and mitigation strategies pertaining specifically to social engineering were prioritized. This process resulted in 210 potentially relevant studies.
3. **Eligibility**: Full-text articles were assessed for eligibility, focusing on their quality, methodological rigor, and relevance to hybrid intelligence frameworks and social engineering attacks. After a detailed evaluation, 90 studies met the inclusion criteria, focusing on primary themes of this literature review.
4. **Inclusion**: Finally, 60 studies were included in the final synthesis, forming the basis for analysis and providing comprehensive insights into hybrid approaches for detecting and combating social engineering attacks.

The PRISMA flow diagram (see Figure 1) provides a visual representation of the review process, highlighting the number of records identified, screened, assessed for eligibility, and included in the final analysis.

PRISMA Flow Diagram

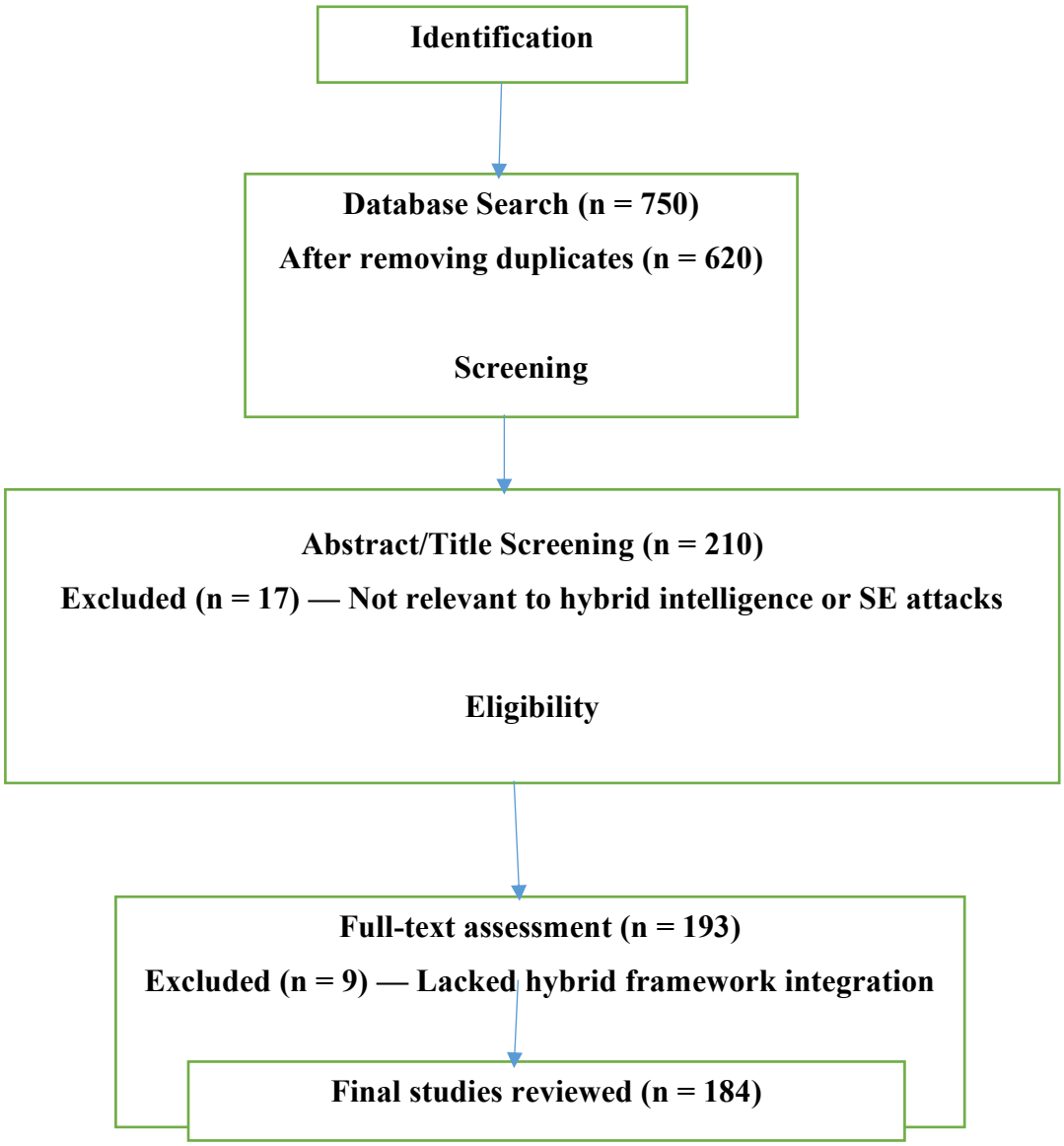

Figure 1: PRISMA Flow Diagram

Review of Literature

The literature can be broadly categorized based on their focus areas: [21] developed a model integrating statistical and machine learning methods for cyber threat detection in IIoT. [22] proposed hybrid optimization strategies. [7] emphasized explainability in AI-based system security. [23] developed a non-linear dynamic model addressing SE, malware, and DDoS. [24] reviewed deep learning for phishing detection. [25] used NLP for SMS spam. [26] proposed adaptive strategies in SEA defense. Game theory was integrated with ML for anticipatory

actions. [2] analyzed modern SE attack types and user susceptibilities. [22]) outlined risks posed by AI-powered impersonation and fraud. [26] studied SE in healthcare. [30] explored deception-based defenses using cognitive AI. These studies reveal how combining human oversight with computational intelligence enhances adaptive capacity in SEA defense frameworks. Table 1 shows a summarized literature findings for the study.

[17] developed an intelligent phishing detection system based on deep learning that automatically learn complex patterns from URL-based features, the model uses a DNN architecture consisting of multiple hidden layers with ReLU activation functions, an output layer with sigmoid activation, and the Adam optimizer for training. The model achieved an accuracy of 98.02%, with a precision of 98.2%, recall of 98%, and F1-score of 98.1%.

[22]. proposed DeepEPhishNet; a deep learning framework for email phishing detection using a combination of word embedding (Word2Vec, FastText, and TF-IDF) and deep learning techniques (DNN and BiLSTM network). The method focuses on using only four header-based features from emails: "From," "Return-path," "Subject," and "Message-ID". The DNN model with FastText-SkipGram achieved an accuracy of 99.52%, while the BiLSTM model with FastText-SkipGram achieved 99.42% showing that DNN achieved slightly better performance.

[6] proposed SEShield, a three-part detection framework designed for in-browser defense against SE campaigns, the first component, SECrawler, is a security crawler that actively collects real-world SE attack data from the web, this data is then used to train SENet, a deep learning-based image classifier capable of identifying distinct visual patterns associated with SE attack pages, the third component, SEGuard, is a proof-of-concept browser extension that integrates SENet for real-time detection as users navigate the internet, their approach reported a detection accuracy of up to 99.6%, with a 1% false positive rate, on new SE attack instances.

[14] proposed a deep learning-based approach to enhance the accuracy of phishing detection and reduce false positives, the study adopts four models: CNN-BLSTM, SNN, Transformer, and DBN, using a phishing dataset that incorporates crucial features like URL structure, domain age, and the presence of HTTPS. Among these models, the CNN-BLSTM model demonstrated the highest accuracy, achieving 98.9%, by effectively linking URL sequences in both space and time.

[9] proposed a phishing detection (PD) browser to address the growing threat of phishing attacks in the context of increased electronic trade and advancements in internet and cloud technologies, the PD browser combines two deep learning algorithms, Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), to create a hybrid model.

[23] introduced the Chat-based Social Engineering Attack Recognition System (CSE-ARS), an innovative defense mechanism against chat-based social engineering (CSE) attacks, the system utilizes a late fusion strategy, combining the outputs of five specialized deep learning models, each model focuses on detecting a specific CSE attack enabler: critical information leakage (CRINL-R), personality traits (PERST-R), dialogue acts (DIACT-R), persuasion (PERSU-R), and persistence (PERSI-R). CSE-ARS employs weighted linear aggregation and simulated annealing with 10-fold cross-validation to optimize its performance, the system was trained on the CSE-ARS Corpus, a dataset specifically designed for CSE attack analysis, evaluations indicate that CSE-ARS effectively identifies and neutralizes CSE threats, thereby enhancing online user security.

Table 1. Summary of Literature Review

Work	Attack Type	Method	ML Technique	Result
Tripathi (2023)	Phishing	Comparative Analysis	LR, DT, RF, AB, KNN, NN, SVM Linear, SVM Poly, SVM rbf, SVM sigmoid, GB, XGB	XGB: 98% Accuracy
Kim (2024)	Phishing	Proactive detection	Logistic Regression	94.72%
He., et al (2023)	Phishing and Insider Threat	Hybrid Model	LSTM and XGBoost	No value reported
Ren., et al (2022)	General SE attacks	Obtained SE threat data from a pre-existing Knowledge Graph (KG)	DT, RF, SVM, MLP, LR, NC, NB, AB, Voting	**DT:** Precision 89.6%, recall 85.5% and F1-score 87.6%

Ishtaiwi et. al, (2024)	Phishing	ML with expert-curated whitelists and blacklists	RF, Tree, NB, LR, SVM, KNN	RF: 97%
Remmide et al, (2024)	General SE attacks	ML with Oversampling	SVM and XGB	99% accuracy
Deep Learning				
Mughahid et al, (2022)	Phishing	Learn complex pattern from URL-based features	DNN	98.02% accuracy
Somesha and Pais (2024),	Phishing	Word Embedding and deep learning techniques	DNN and BiLSTM network	accuracy of 99.52%
Ozen et al, (2024)	All SE attacks	three-part detection framework	SECrawler, SENet & SEGuard	accuracy of up to 99.6%,

Discussion

This review identifies the prevailing trends in Hybrid Intelligence Systems: Tools integrating decision support with user intervention show more flexibility and context-awareness. Studies such as [22] emphasized interpretability, especially in regulated environments. Real-world SEA datasets remain scarce. Synthetic datasets, though helpful, lack behavioral realism. Many hybrid models are validated in controlled environments, limiting field deployment. Integrating behavioral science, cognitive psychology, and AI engineering is necessary to develop realistic hybrid intelligence systems.

Research Gaps

Key limitations and gaps identified across the literature include:

i. Lack of standardization across frameworks.

ii. Scarcity of real-world datasets that mimic complex SE attacks.

iii. Difficulty in quantifying human intelligence contribution within hybrid models.

iv. Limited comparative benchmarking of hybrid vs. pure-AI models.

v. Fragmentation in simulation and validation methods.

Findings and Discussion

The systematic review results categorized the literature into three primary themes: detection mechanisms, user behavior and awareness, and frameworks for prevention.

Detection Mechanisms

Detection mechanisms involve identifying potential social engineering attacks using various technologies. The integration of machine learning within hybrid intelligence frameworks significantly enhances detection capabilities.

Machine Learning Algorithms in Detection Machine learning has garnered significant focus in social engineering attack detection due to its capability to process vast amounts of data and identify patterns that are often imperceptible to human analysts. Various studies have illustrated that models like Random Forest, Support Vector Machines (SVM), and Neural Networks can effectively classify benign from malicious interactions [26][8].

An example is the work of Mouton et al., which introduced the SEADM (Social Engineering Attack Detection Model), employing a finite state machine approach to systematically profile social engineering attack interactions. This model incorporates machine learning algorithms to improve classification accuracy in detecting different types of social engineering attacks such as phishing, vishing, and pretexting [1].ML techniques train algorithms on datasets of legitimate and phishing websites to learn distinguishing patterns. These methods adapt to evolving phishing techniques and handle large volumes of data. However, effective training and feature selection are crucial for accuracy, and regular updates are needed as phishing tactics change [11].

[12] conducted a comparative study on the use of supervised machine learning models for phishing website detection. they implemented and tested 12 different classifiers. The results demonstrated that ensemble methods outperformed other classifiers, with XGBoost achieving the highest

accuracy of 98.32% closely followed by random forest and neural networks. Kim (2024), designed a phishing detection model that predicts the likelihood of a phishing attempt using machine learning, the core aim was to move beyond reactive countermeasures to a proactive detection system, sklearn logistic regression was used to train the model and it achieved an accuracy of 94.72%. [19] proposed a hybrid model combining Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) to detect phishing emails, while the **Group Security Layer** uses a Bidirectional LSTM (Bi-LSTM) integrated with an Attention mechanism to identify insider threats. Their method achieved an high accuracy but the exact value was not reported.

[40] explores the use of ML techniques for detecting general Social Engineering attacks, they processed and obtained SE threat data from a pre-existing Knowledge Graph (KG), extracted various threat features and created new datasets based on three different feature combinations, nine ML algorithms were trained and tested with **Decision Tree model,** achieving the best performance with average precision, recall and F1-score of 89.6%, 85.5% and 87.6% respectively. [25] proposes a novel framework integrating ML with expert-curated whitelists and blacklists to advance anti-phishing efforts, six ML algorithms were compared, and the RF classifier achieved over 97% accuracy, which demonstrates the approach's effectiveness for accurately detecting phishing websites and the benefits of combining data-driven learning with human expertise.

[18] proposes a novel machine learning approach to improve the detection of social engineering attacks, The researchers analyzed support vector machines (SVM) and XGBoost models on imbalanced datasets and utilized oversampling techniques, specifically SMOTE-ENN, to handle class imbalance. Their experimental results indicate that the oversampled SVM model achieved over 99% accuracy in attack detection, outperforming other methods. A statistical analysis using ANOVA confirmed a significant improvement in detection performance compared to previous approaches.

Behavioral Analysis and Contextual Awareness: In addition to traditional machine learning methods, the integration of behavioral analysis into detection mechanisms has emerged as a promising approach. Hybrid frameworks can enhance the efficacy of detection systems by leveraging social engineering psychology principles to identify behavioral anomalies indicative of an ongoing attack. This method emphasizes understanding the psychological triggers exploited by attackers, which can provide additional context in the analysis of user interactions.

For instance, the research conducted by Wang et al. introduced a structured ontology for social engineering attacks, facilitating better understanding and classification of attack vectors within the cybersecurity domain [6].This ontology provides a framework through which machine learning models can contextualize user actions and identify potential security breaches more accurately.

User Behavior and Awareness

The criticality of user behavior is underscored in much of the literature, emphasizing the need for organizational strategies aimed at enhancing user awareness of social engineering tactics.

The Role of User Training and Awareness

[2]. highlighted that user awareness and education are paramount in changing employee behavior—an essential line of defense that organizations must bolster [3] . Effective training programs that simulate social engineering tactics can dramatically decrease vulnerability. Studies have shown that awareness campaigns—especially those involving realistic simulations of phishing attacks—significantly improve employees' response to suspicious communications ([15]

Psychological Factors in Social Engineering

Understanding the psychological triggers that social engineering attacks exploit is crucial for training users effectively. Studies indicate that factors such as fear, urgency, and authority play significant roles in how individuals react to social engineering ploys. Consequently, training sessions should provide information on these social engineering tactics and reinforce the importance of skepticism regarding unsolicited communications [21], [24].

Culture of Security Awareness :The literature consistently advocates for fostering a culture of security awareness within organizations. A participatory culture that encourages employees to report suspicious activity without fear of judgment can enhance overall organizational resilience to social engineering attacks. Initiatives fostering open communication create an environment where security becomes a collective responsibility rather than an isolated IT function [3]; [14]

Frameworks for Prevention

Proposed frameworks exemplify methodical approaches to address social engineering threats comprehensively.

Security Pattern-Based Analysis Frameworks

Research by [21]. proposed a security pattern-based analysis framework that systematically generates countermeasures based on the breakdown of social engineering tactics. This framework identifies common attack vectors and provides preventive measures tailored to these threats, facilitating organizations in developing proactive defense strategies against social engineering [5].

Comprehensive Approaches Incorporating Human and Technological Elements

[47]. explored comprehensive approaches that combine both technological defenses and human factors in mitigating social engineering attacks. Their research indicates that a dual focus on technology and human behavior results in a more robust organizational response to potential threats [26]. For instance, hybrid frameworks could incorporate automated systems for early detection while concurrently undertaking user training to raise awareness and reduce susceptibility to attacks.

Evaluation of Framework Effectiveness Future studies should focus on evaluating the effectiveness of these proposed frameworks in real-world settings. Implementing pilot programs that measure improvements in detection rates, incident reporting, and changes in employee behavior can enhance understanding of what strategies are most effective in combating social engineering attacks. By quantifying results, organizations may be better equipped to justify investments in security measures based on demonstrated effectiveness.

Challenges and Limitations of Existing Research

While the review identifies promising approaches and frameworks, notable challenges persist in the field of detecting and mitigating social engineering attacks.

1. **Rapidly Evolving Threat Landscape** :The fast-paced evolution of social engineering tactics poses challenges for machine learning models that require training on current data. Studies often rely on datasets that may not reflect the most recent attack methodologies, leading to potential gaps in detection capabilities [13].
2. **Integration of Diverse Data Sources** :Hybrid intelligence frameworks often depend on multiple sources of data for effective analysis. Integrating diverse data types—such as employee behavior, historical incident reports, and external threat intelligence—remains a complex task that not all organizations can proficiently manage [25]. Comprehensive vulnerability assessments must take a holistic approach to data aggregation for context-driven insights.

3. **User Resistance to Awareness Programs** User training and awareness programs can face resistance; employees may feel burdened by additional training or skeptical about the likelihood of social engineering attacks impacting them. Addressing this resistance requires not only robust training content but also an understanding of organizational dynamics and employee perspectives on cybersecurity [10].
4. **Resource Allocation for Implementation** For many organizations—particularly smaller enterprises—allocating resources for training, advanced technologies, and comprehensive detection frameworks can be a substantial hurdle. Budget constraints may limit the ability to implement expansive security measures, an issue that research must continue to explore [2].

Future Research Directions

This systematic literature review highlights critical areas for future research to develop hybrid intelligence frameworks aimed at combating social engineering attacks.

1. **Longitudinal Studies on Effectiveness of Training Programs:** Future studies should employ longitudinal methodologies to evaluate the long-term effectiveness of user education and training programs on susceptibility to social engineering attacks. Measuring specific behavioral changes and incident reporting over time can provide valuable insights into the sustainable impact of training initiatives.
2. **Integration of Advanced Technologies:** Investigating the integration of emerging technologies—such as generative AI and natural language processing—within hybrid frameworks offers rich potential for enhancing detection and mitigation as these technologies continue to evolve [4]. Experimental applications in real-world scenarios could provide insights into practical deployments.
3. **Behavioral Analytics and Predictive Modeling:** Research aimed at enhancing predictive modeling based on behavioral analytics could yield more accurate insights into potential social engineering attack trajectories. By examining patterns of legitimate versus deceptive communication, organizations can anticipate threats more effectively and apply preventive measures preemptively.
4. **Creation of Standardized Metrics for Evaluation:** Developing standardized metrics for evaluating the effectiveness of hybrid intelligence frameworks will aid organizations in measuring improvement and justifying investments in upgraded security measures.

Tailoring metrics to different organizational contexts may enhance their relevance and make them broadly applicable.

5. **Investigating the Role of Leadership in Security Culture** :Understanding how leadership approaches security and the role it plays in shaping organizational culture could provide a basis for more effective security initiatives. Future research should explore the influence of managerial attitudes towards cybersecurity and their impact on employee engagement and behavior.
6. **Cultural Considerations in Awareness Programs** Further research should account for cultural differences in behavioral responses to cybersecurity threats. Understanding how individual behaviors change across different cultural settings can inform the design of more effective training programs that resonate with diverse workforces[26].

Conclusion

The exploration of hybrid intelligence frameworks presents promising avenues for bolstering defenses against social engineering attacks. By synthesizing advanced technological measures with psychological principles, organizations can cultivate a more resilient cybersecurity posture. This systematic literature review underscores the importance of user behavior, detection mechanisms, and comprehensive frameworks in mitigating risks associated with social engineering.

By providing insights into the existing body of knowledge and identifying critical trends, gaps, and avenues for future research, this review emphasizes the need for a multidisciplinary perspective that combines behavioral and technological insights. As organizations continue to navigate the complexities of an increasingly digital landscape, adopting hybrid intelligence frameworks will be essential in reinforcing frameworks to confront and fortify against the pervasive threat of social engineering.

References

1. Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in Internet of Things using artificial intelligence methods: A systematic literature review. *Electronics, 11*(2), 198. https://doi.org/10.3390/electronics11020198.

2. Matyokurehwa, P., & Mhlanga, D. (2022). Enhanced social engineering framework mitigating against social engineering attacks in higher education. *Security and Privacy*, 5(4). doi:10.1002/spy2.237.
3. Alshaikh, M. (2022). Exploring the effectiveness of cybersecurity policies in organizations: A case of employee sentiment and resistance. *Information Management & Computer Security*, 30(2), 109-125.
4. Kaur, B. (2024). Social engineering attacks in the digital age. *International Journal of Social Science and Education (IJSSE).*
5. Schlegel, R. (2021). The importance of a security culture in cybersecurity: A framework for organizations. *Journal of Information Security and Applications*, 56, 102658. doi:10.1016/j.jisa.2020.102658.
6. Wright, J. (2023). Assessing the cost-effectiveness of social engineering mitigation strategies in organizations. *Network Security*, 2023(11), 37-45. doi:10.1016/j.netsec.2023.10.005.
7. Wen, S.-F., Shukla, A., & Katt, B. (2024). Artificial intelligence for system security assurance: A systematic literature review. *International Journal of Information Security, 24*(1), 43. https://doi.org/10.1007/s10207-024-00959-0
8. Mitnick, K. D. (2014). The Art of Deception: Controlling the Human Element of Security. Wiley.
9. Goorha, P. (2019). Employee perceptions of security awareness programs: Implications for security policy in organizations. *Cybersecurity Review*, 3(2), 14-25.
10. Azevedo, B. F., Rocha, A. M. A. C., & Pereira, A. I. (2024). Hybrid approaches to optimization and machine learning methods: A systematic literature review. *Machine Learning, 113*(7), 4055–4097. https://doi.org/10.1007/s10994-023-06467-x
11. Schmitt, K., & Fléchais, I. (2024). Digital deception: generative artificial intelligence in social engineering and phishing. *Artificial Intelligence Review*. doi:10.1007/s10462-024-10973-2
12. Prasad, N., & Koller, S. (2022). The psychological manipulation of security awareness: Context's role in shaping outcomes. *International Journal of Cybersecurity Intelligence & Cybercrime*, 5(1), 39-52.

13. Mensah, J., & Yu, Y. (2023). Cultural dimensions of cybersecurity awareness in multinational organizations: An analytical framework. *Journal of Risk Research*, 26(2), 145-169.
14. Yu, J., Yu, Y., Wang, X., Lin, Y., Yang, M., Qiao, Y., & Wang, F.-Y. (2024). The shadow of fraud: The emerging danger of AI-powered social engineering and its possible cure. *arXiv*. http://arxiv.org/abs/2407.15912
15. Mohan, P. V., Dixit, S., Gyaneshwar, A., Chadha, U., Srinivasan, K., & Seo, J. T. (2022). Leveraging computational intelligence techniques for defensive deception: A review, recent advances, open problems and future directions. *Sensors, 22*(6), 2194. https://doi.org/10.3390/s22062194
16. Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: a framework for understanding effective cyber security. *Journal of Information Systems Security*, 5(3), 10-22.
17. Syafitri, R., & Rahmawati, G. (2022). Social Engineering Attacks Prevention: A Systematic Literature Review. *Ieee Access*, 10, 9451-9461. doi:10.1109/access.2022.3162594
18. Jing, J. (2025). Applications of game theory and advanced machine learning methods for adaptive cyberdefense strategies in the digital music industry. *Future Internet.*
19. Benavides, E., Fuertes, W., Sanchez, S., & Sanchez, M. (2020). Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review. In A. Rocha & R. P. Pereira (Eds.), *Developments and Advances in Defense and Security* (pp. 51–64). Springer. https://doi.org/10.1007/978-981-13-9155-2_5
20. Bouke, M. A., & Abdullah, A. (2024). SMRD: A novel cyber warfare modeling framework for social engineering, malware, ransomware, and distributed denial-of-service based on a system of nonlinear differential equations. *Journal of Applied Artificial Intelligence, 5*(1). https://doi.org/10.48185/jaai.v5i1.972
21. Saidat, M. R. A., Yerima, S. Y., & Shaalan, K. (2024). Advancements of SMS spam detection: A comprehensive survey of NLP and ML techniques. *Procedia Computer Science, 244*, 248–259. https://doi.org/10.1016/j.procs.2024.10.198

22. Mouton, F., & Hsiao, K. (2018). Finite State Machine for the Social Engineering Attack Detection Model: SEADM. *Saiee Africa Research Journal*, 1084-1091. doi:10.23919/saiee.2018.8531953
23. Wang, T., Yang, Y., & Zhang, C. (2021). Social engineering in cybersecurity: a domain ontology and knowledge graph application examples. *Cybersecurity*, 4(1), tyab004. doi:10.1093/cybersec/tyab004
24. Li, A., Zhang, T., & Zhao, L. (2023). Defending against social engineering attacks: A security pattern-based analysis framework. *Iet Information Security*, 17(3), 275-288. doi:10.1049/ise2.12125
25. Nifakos, S., Chandramouli, K., Nikolaou, C. K., Papachristou, P., Koch, S., Panaousis, E., & Bonacina, S. (2021). Influence of human factors on cyber security within healthcare organisations: A systematic review. *Sensors, 21*(15), 5119. https://doi.org/10.3390/s21155119
26. Liu, Y., Li, S., Wang, X., & Xu, L. (2024). A review of hybrid cyber threats modelling and detection using artificial intelligence in IIoT. *Computer Modeling in Engineering & Sciences, 140*(2), 1233–1261. https://doi.org/10.32604/cmes.2024.046473